Data Science Workflow

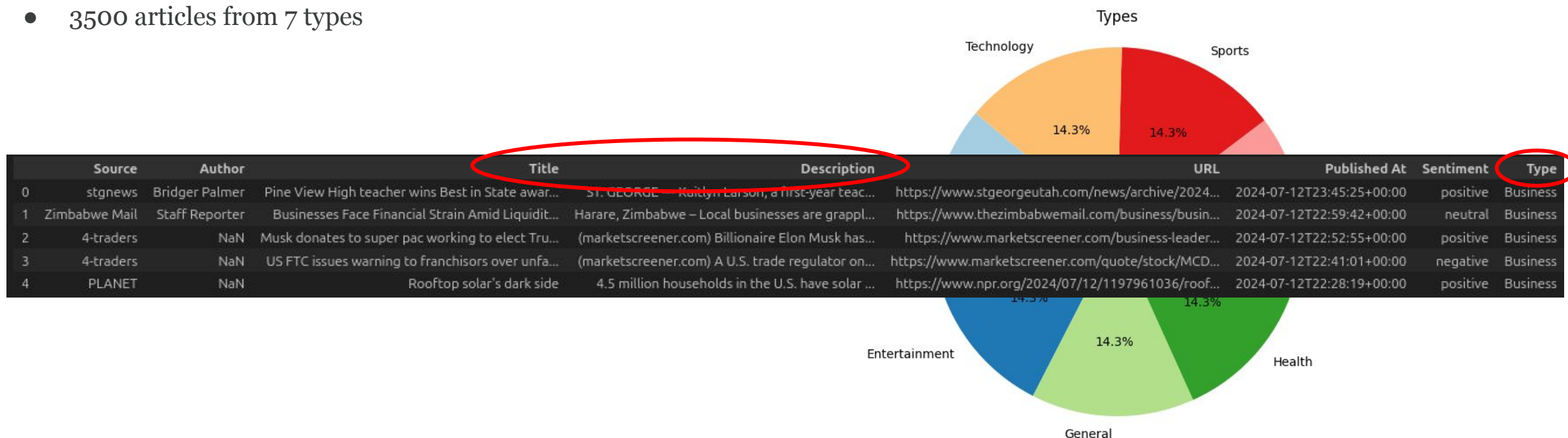## News Aggregator

10.08.2024
Nicolas Hahn

Content

1. News Aggregator

- Sources: Al Jazeera, CNN, Fox News, New York Times

- Receive messages from sections: news, economy, politics, business, technology

- Multiple days while ensuring that no duplicate news

- New York times uses JSON script, different methods needed

- Create data frame and csv: articles.csv

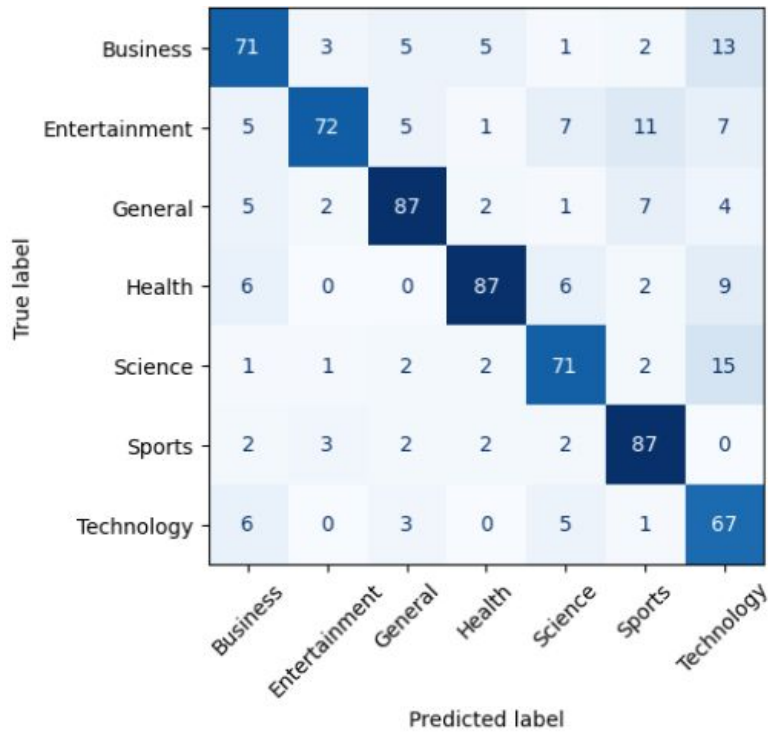|	Title	Summary	Date	Author	Source	URL	Title Length	Summary Length	Language
0	Kerala landslides death toll hits 151, many missing as rain hampers rescue	Hundreds of rescue workers search through mud and debris for a second day after multiple landslides hit a hilly district in India's southern state.\nHundreds of rescue workers are searching through mud and debris from multiple landslides that have killed at least 151 people in southern India's Kerala state, police said.\n"The landslides happened at the top and then affected the lower valley areas where people were staying.\nTelevision footage showed rescue workers making their way through mud and uprooted trees to reach those who had been stranded.\nNearly 500 people were killed around Kerala in 2018 during the worst flooding to hit the state in almost a century.	07/31/2024	NaN	www.aljazeera.com	https://www.aljazeera.com/news/2024/7/31/india-kerala-landslides-death-toll-hits-151-many-missing-as-rain-hampers-rescue	74	667	en
1	World reacts to killing of Hamas political chief Haniyeh in Iran	Hamas's political chief was in Tehran when he was killed, with the Palestinian group that governs Gaza blaming Israel.\nHamas's political leader Ismail Haniyeh has been killed in Iran's capital Tehran.\nHis death has been confirmed in separate statements by the Palestinian group that runs the Gaza Strip and Iran's elite Islamic Revolutionary Guard Corps.\nIsrael is on the verge of collapse, and its reactions reflect confusion and inability to achieve any of its goals.\nIranian Foreign Ministry spokesman Nasser Kanaani"Haniyeh's martyrdom in Tehran will strengthen the deep and unbreakable bond between Tehran, Palestine, and the resistance," Kanaani said, according to Iranian state media.	07/31/2024	NaN	www.aljazeera.com	https://www.aljazeera.com/news/2024/7/31/reactions-to-the-killing-of-hamass-ismail-haniyeh	64	691	en

2. Classification

- Dataset from kaggle: https://www.kaggle.com/datasets/clovisdalmolinvieira/news-sentiment-analysis/code

- 3500 articles from 7 types
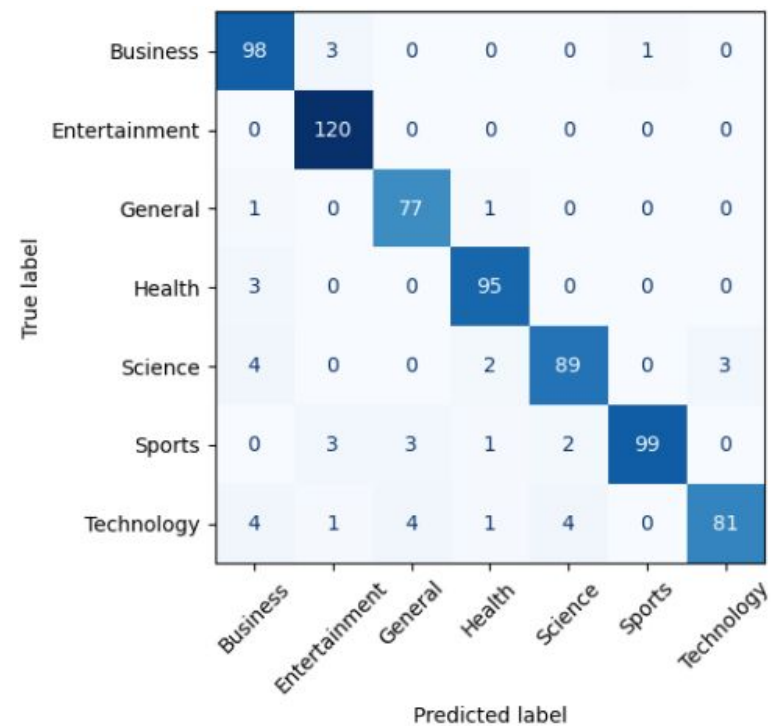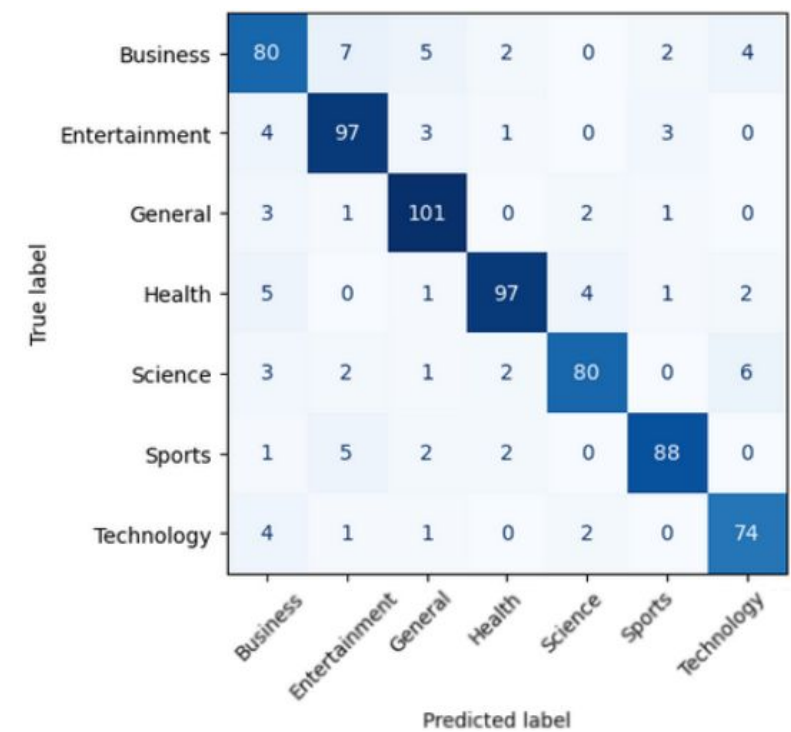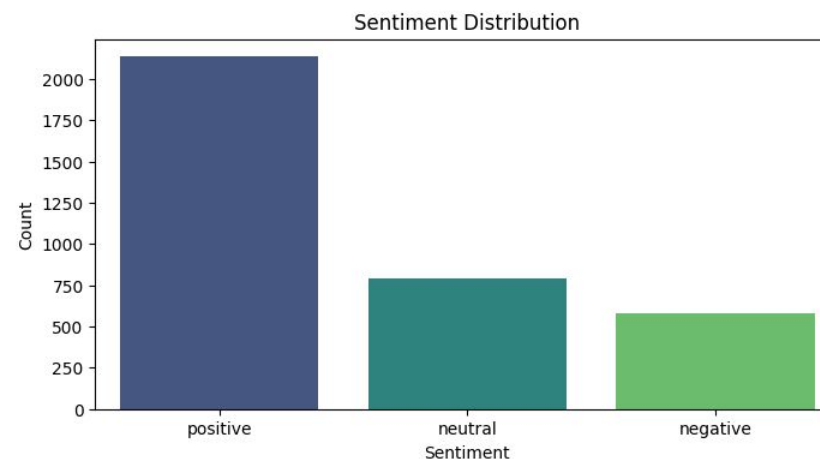
2. Classification
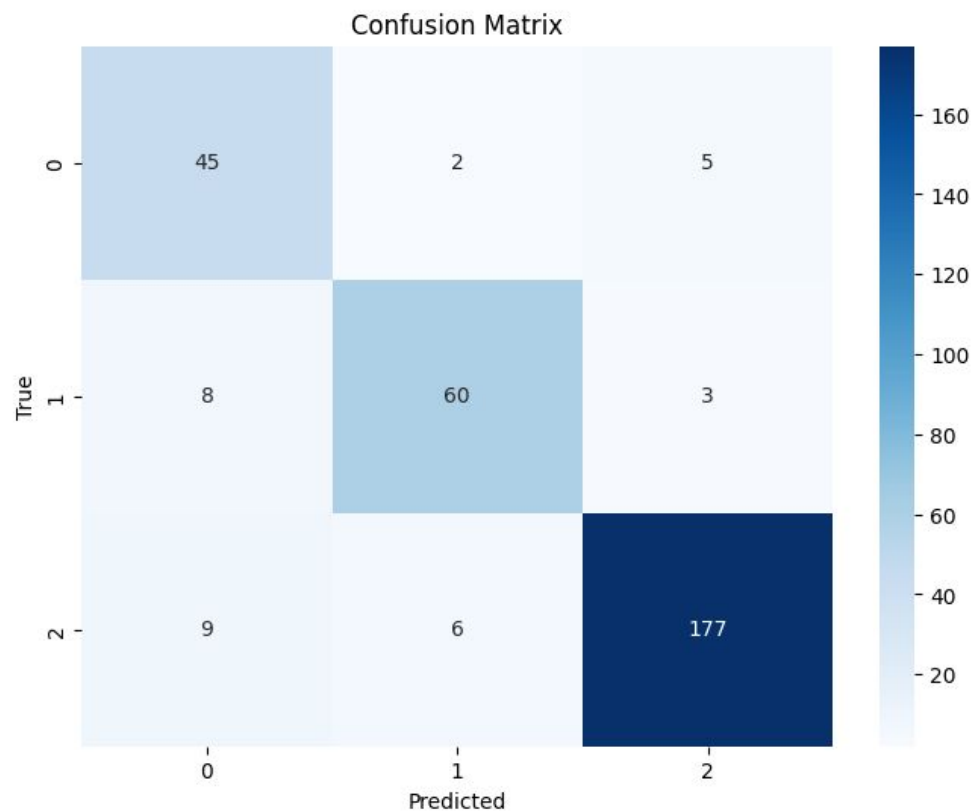

NaiveBayes   Bert   Autogluon-LightGBM

3. Sentiment

- Sentiment: positive=2, neutral=1, negative=0

- Bert model: bert-base-uncased, 5 epochs

- Predict sentiment with description

- Use bert tokenizer

Sentiment Distribution

	Source	Author	Title	Description	URL	Published At	Sentiment	Type
0	stgnews	Bridger Palmer	Pine View High teacher wins Best in State awar...	ST. GEORGE — Kaitlyn Larson, a first-year teac...	https://www.stgeorgeutah.com/news/archive/2024...	2024-07-12T23:45:25+00:00	positive	Business
1	Zimbabwe Mail	Staff Reporter	Businesses Face Financial Strain Amid Liquidit...	Harare, Zimbabwe – Local businesses are grappl...	https://www.thezimbabwemail.com/business/busin...	2024-07-12T22:59:42+00:00	neutral	Business
2	4-traders	NaN	Musk donates to super pac working to elect Tru...	(marketscreener.com) Billionaire Elon Musk has...	https://www.marketscreener.com/business-leader...	2024-07-12T22:52:55+00:00	positive	Business
3	4-traders	NaN	US FTC issues warning to franchisors over unfa...	(marketscreener.com) A U.S. trade regulator on...	https://www.marketscreener.com/quote/stock/MCD...	2024-07-12T22:41:01+00:00	negative	Business
4	PLANET	NaN	Rooftop solar's dark side	4.5 million households in the U.S. have solar ...	https://www.npr.org/2024/07/12/1197961036/roof...	2024-07-12T22:28:19+00:00	positive	Business

3. Sentiment

Confusion Matrix

	precision	recall	f1-score	support
Class 0	0.73	0.87	0.79	52
Class 1	0.88	0.85	0.86	71
Class 2	0.96	0.92	0.94	192
accuracy			0.90	315
macro avg	0.85	0.88	0.86	315
weighted avg	0.90	0.90	0.90	315

4. Results

- Try Bert model on own news dataset:

Title: Mali cuts diplomatic ties with Ukraine over Wagner attack controversy

Summary: Move follows Kyiv's official's comments on attacks late last month targeting Malian soldiers and mercenaries from Russia's Wagner group.
Mali has announced it is severing diplomatic ties with Ukraine, accusing a senior official of having admitted Kyiv's role in a heavy defeat in July that killed dozens of mercenaries from Russia's Wagner group as well as Malian soldiers.
Ukraine's actions violated Malian sovereignty and constituted unacceptable foreign interference and support for international terrorism, according to Maiga.
Mali's army has admitted it suffered a "large number" of deaths during the clashes but has not released figures.
They have also moved closer to Russia, and Wagner has been operating in Mali since late 2021, replacing French troops and international peacekeepers.

Predicted Sentiment: 0

Title: How YouTube Took Over Our Television Screens

Summary: Two years ago, YouTube abandoned its audacious plan to beat Hollywood at its own game.
So it canceled the shows and fell back on the user-created content that had made it a household name.
Now, YouTube consistently ranks as the most popular streaming service on U.S. televisions, surpassing the companies it once tried to emulate.
Since Netflix started offering original content in 2012, TV networks have jumped into the streaming race, trying to outflank one another with major upfront investments.
Netflix alone spends $17 billion a year on new series and movies as well as on older fare from other companies' libraries.
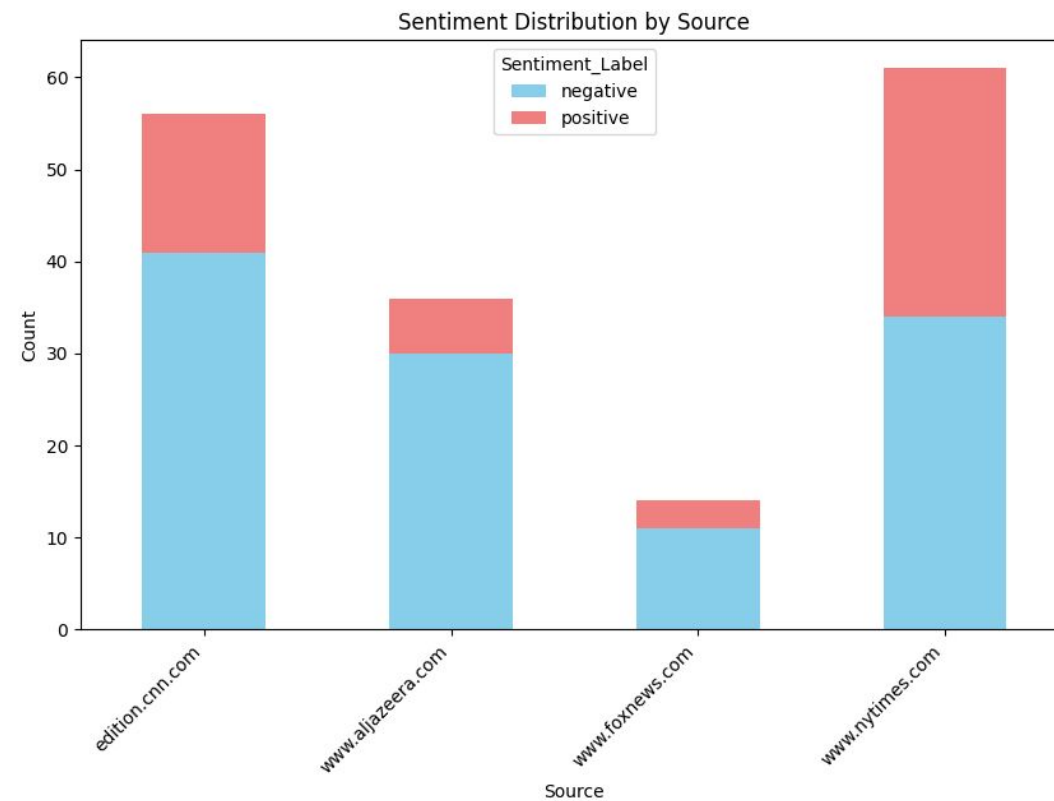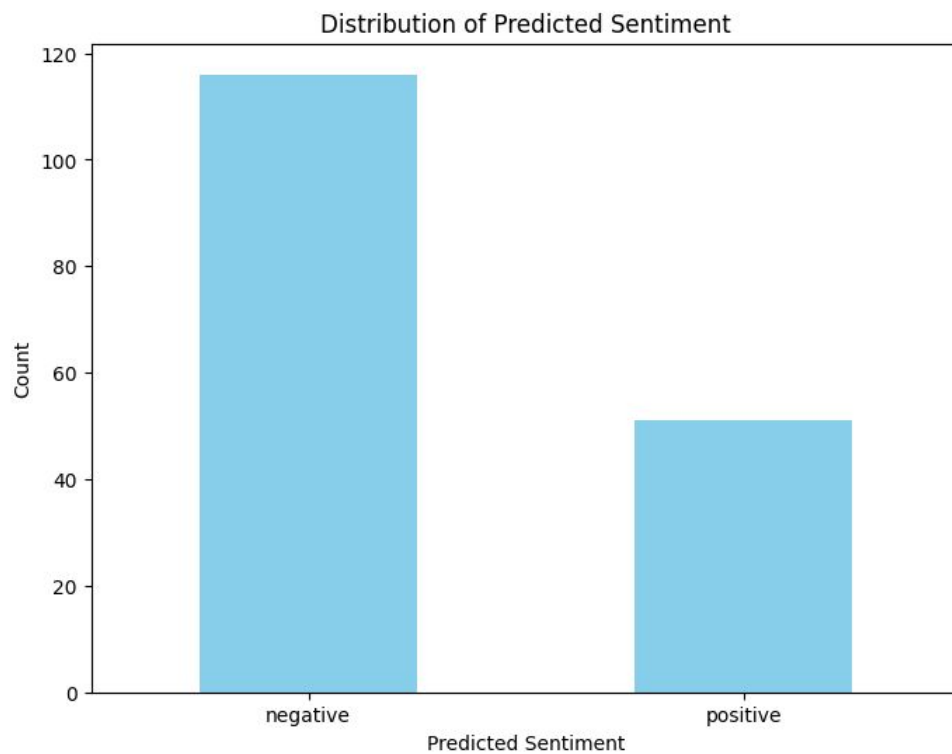
Predicted Sentiment: 2

Title: How Kamala Harris' candidacy has energized plans for the DNC

Summary: Officials quickly moved to tailor the Chicago convention to Harris with only four weeks, in part drawing from the 2008 convention when Barack Obama accepted the nomination given the historic candidacy of Harris herself.
Nominee acceptance speeches for the convention, which begins the week of August 19, are currently scheduled for that Wednesday and Thursday night, according to convention officials.
Hollywood gets on boardThe Hollywood community is clamoring to support Harris at the DNC, according to numerous agents, publicists and Democratic strategists who spoke to CNN.
No performers have been set yet, though various representatives for musicians have been in talks with the Harris campaign, according to two sources.
For instance, besides Megan Thee Stallion's performance at her Atlanta rally earlier this week, Harris had the support of the rapper Quavo, who spoke about gun violence prevention.

Predicted Sentiment: 0

4. Results

Source

1. https://medium.com/coinmonks/how-to-implement-a-recommendation-system-with-deep-learning-and-pytorch-2d40476590f9
2. https://www.kaggle.com/code/emirkocak/in-depth-series-sentiment-analysis-w-transformers
3. https://github.com/dani2221/MediaBiasClassifier
4. https://medium.com/@danilo.najkov/detecting-political-bias-in-online-articles-using-nlp-and-classification-models-c1a40ec3989b
5. https://medium.com/@jonathanmondaut/building-a-news-aggregator-with-python-and-beautiful-soup-da7784f9136a
6. https://huggingface.co/docs/transformers/index
7. https://github.com/Rohak72/Newspaper-Scrape
8. https://huggingface.co/docs/transformers/model_doc/bert
9. https://arxiv.org/abs/1810.04805

Thanks